chinaXiv:201908.00125v1

脑功能成像数据的标准化存储框架（BIDS）和 BIDS APP 分析流程探讨

陈圣栋，高伟，罗利，杨洁敏，袁加锦

（西南大学心理学部，认知与人格教育部重点实验室，重庆， 400715）

通信作者:袁加锦, E-mail: yuanjiaj@swu.edu.cn; yuanjiajin168@126.com

中国自然科学基金(NSFC31671164, NSFC31871103)和中央高校基本科研业务费专项资金资助(SWU1809359)

摘要 无损的功能性磁共振脑成像（functional magnetic resonance imaging，fMRI）技术已经成为脑科学研究领域的重要研究手段之一。然而，全球不同实验室在 fMRI 数据的存储和处理流程上存在一定的分歧，限制了脑科学的快速发展。为了解决这一问题，最近国际脑科学家联合推出了脑成像数据存储的标准化框架（Brain Imaging Data Structure, BIDS），以及能够处理以 BIDS 格式整理的数据的应用程序（BIDS APP）。本文首先概述了 BIDS 以及 BIDS APP 的处理流程，然后讨论了由于对 BIDS 数据预处理技术的选择不同所导致的不同的后续统计分析流程。面对新兴的 BIDS 程序和技术，未来的脑成像研究者需要考虑如何将其与传统的 fMRI 统计软件合理搭配以期达到更好的统计力度。

关键词：脑成像，功能性磁共振，数据分析流程，BIDS，BIDS APP

1. 前言

磁共振脑成像技术在探究人脑结构、功能以及各种与脑相关的疾病诊断中正起着越来越重要的作用。以功能性磁共振成像 (functional magnetic resonance imaging，fMRI)为关键词，在科学基金共享网对已经结题的国内的科研项目进行检索，结果发现一共有 323 项认知神经科学或医学相关的相关项目，总资助金额达数千万人民币。这表明 fMRI 技术已经成为国内脑科学相关

chinaXiv:201908.00125v1

学科的重要研究手段。

然而，不同研究机构乃至同一个机构的不同课题组由于各种各样的原因，在脑成像数据的存放、整理以及处理的方法上存在较大的差别。由于脑成像数据的处理需要一定的编程能力，这种混乱和孤立的情况造成不同的研究者需要花费大量的时间和精力浪费在数据处理代码的编写和纠错上。另一方面，近期研究表明通过一些新的预处理技术能够有效降低功能性 MRI 数据中噪声（如呼吸、心跳和头动噪声），进而提高脑成像统计分析的效力(Behzadi, Restom, Liau, & Liu, 2007; Pruim et al., 2015)。但是，这些新的预处理技术本身较为复杂，学习成本较高，种种因素直接导致国内的研究者对其接受度不高，难以应用到具体研究中。为了解决这两个方面的问题，2016 年由斯坦福大学牵头，剑桥、牛津等多所大学脑成像中心联合刊文，推出了一个标准化的脑成像数据框架（Brain Imaging Data Structure, BIDS）(Gorgolewski et al., 2016)（**图 1**）。BIDS 的推出意味着全世界顶级脑科学家在脑成像数据的整理和存储方面达成了初步的共识。紧接着，在 2017 年，这些脑科学家又推出了一系列能够处理以 BIDS 格式整理的数据的应用程序（BIDS APPs）(Gorgolewski et al., 2017)。但是，脑成像数据的处理一般都需要几个阶段，如预处理和统计分析（包括个体分析和组分析）；而目前的 BIDS APPs 往往只针对其中某一个或几个阶段的数据分析。例如斯坦福大学开发的 MRIQC (Esteban et al., 2017)这一软件专注于对采集的原始数据进行质量评估和控制，而 fmriprep 这一 BIDS APP 专注于对通过质量筛查的数据进行预处理。虽然预处理阶段 BIDS APP 的整合使用非常清楚（先 MRIQC 再 fmriprep），但是目前如何从预处理 BIDS APP 过渡到传统的统计分析的工具包（如 SPM），仍然令人困惑。

而且，近几年一些新提出的脑成像数据预处理技术让从预处理 BIDS APP 过渡到传统的统计分析的工具包这一过程变得更加复杂。新兴的预处理技术中具有代表性的分别为将主成份分析和独立成分分析应用到预处理中的 ComCor 和 ICA-AROMA 方法。ComCor 主要用以分离和提取生理噪声的相关信号(Behzadi et al., 2007)，而 ICA-AROMA 主要用以分离和去除头动噪声的相关信号(Pruim et al., 2015)。在血氧水平依赖（blood-oxygen-level dependent, BOLD）和动脉自旋标记

(Arterial Spin Labelling, ASL)两种静息态 fMRI 像数据分析中，ComCor 相比传统的回归分析方法显著降低了其数据的时间标准偏差；而且 ComCor 还能提高 BOLD 和 ASL 两种任务态（视觉加工任务）脑成像数据的统计功效。类似地，ICA-AROMA 技术在静息态和任务态中的应用同样比传统的回归方法更有效地去除了头动对数据的干扰。但是，由于这些方法提出的时间距今较短，如何在预处理中合理地选择和使用这些新方法，以及如何合理地统筹预处理和统计分析的处理步骤，目前没有可靠的经验可以参考。这直接限制了新技术的应用和推广。

为了解决上诉问题，本文对基于 BIDS 数据预处理过程中两种新兴的预处理技术的应用，以及后续的统计分析流程进行了概述和讨论，归纳整理了合理的可行的分析流程，试图为国内的脑成像工作者提供的一定的参考依据。本研究把基于 BIDS 的数据分析流程初步定义为四步，下面简要阐述每一步骤。

2. BIDS APP 数据预处理和后续统计分析流程探讨

在做完 fMRI 实验并拿到原始的磁共振数据之后，脑成像研究者需要做的第一步工作就是对原始 fMRI 数据的存储格式进行了整理，使其组织框架符合 BIDS 的标准（**图 1**）。使 fMRI 数据的存储框架符合 BIDS 标准是使用后续 MRIQC 以及 fmriprep 等 BIDS APP 的前提。这样只需要将 BIDS 目录作为 BIDS APP 的输入目录，以及设置一些简单的参数，BIDS APP 就会自动化地对输入的数据进行分析。

chinaXiv:201908.00125v1

chinaXiv:201908.00125v1

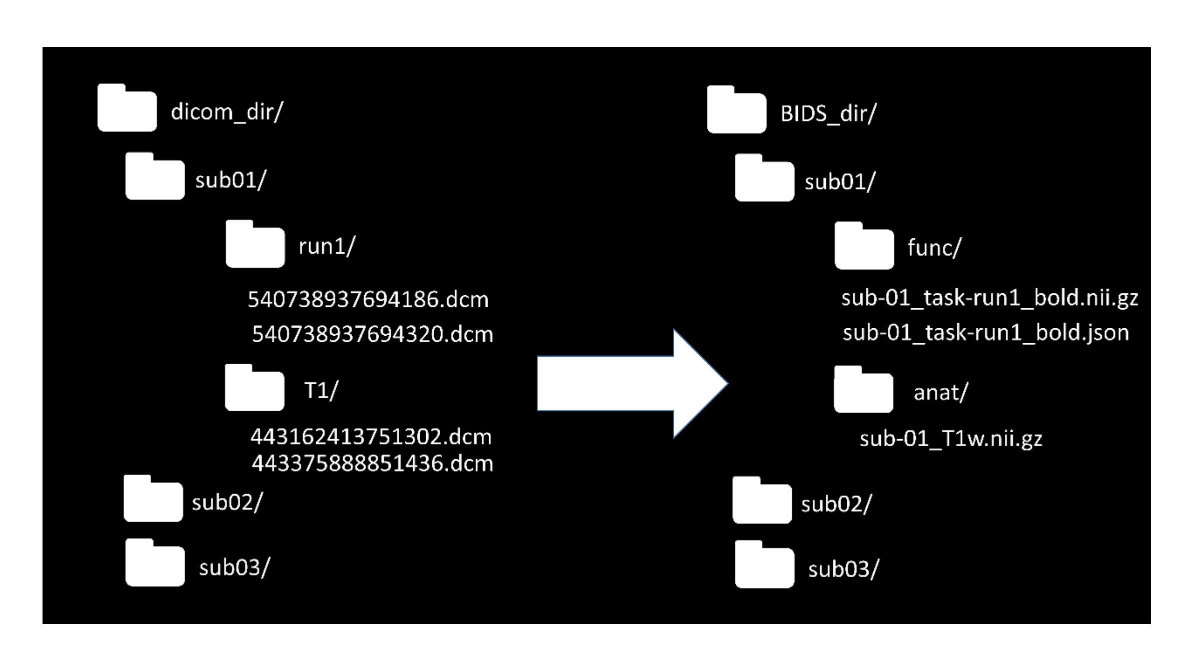


图 1. 西门子 MRI 机器生成的原始 dicom 格式数据存储格式和转换之后的 nii 格式的 BIDS 数据存储格式。

第二步，研究者需要使用 MRIQC 工具包对数据的质量情况进行检查，剔除质量不合格的被试数据。MRIQC 和 fmriprep 一样是一款基于 BIDS 标准的 APP，将基于 BIDS 格式的数据作为输入，可以得到每个被试（个体水平）以及所有被试（组水平）汇总的质量检测结果。MRIQC 提供了丰富的质量控制指标，但由于本文主要关注预处理和统计分析的过程，故不做过多的描述。目前研究者主要采用了其中的逐帧位移（Framewise Displacement， FD）指标作为质量筛选标准。FD 表示逐层扫描时，层与层之间的位移情况，也就是即时头动的情况。MRIQC 报告的 FD 是一个 session 过程中的平均值，称之为“fd_mean”。根据前人文献，建议剔除了 FD（fd_mean）大于 0.2 mm 的被试数据 (Power, Barnes, Snyder, Schlaggar, & Petersen, 2013)。目前开发者对前两步的操作目前没有争议，所以我们建议所有的数据先进行这两步的处理（数据整理+质量控制）。

第三步，使用 fmriprep 对数据进行预处理（**图 2**）。此时，研究者就需要考虑预处理技术的选择问题。预处理过程中主要需要考虑的问题是是否要对数据进行 ICA-AROMA 分析，其他的预处理过程和传统的预处理方法基本一致，不做过多赘述。预处理过程中，ComCor 和 ICA-AROMA 是在 BOLD 数据空间标准化操作完成后分别独立进行的，两种方法之间互不影响。但两

种方法对数据的处理方式上差别很大，ComCor 只提取功能像数据中与生理噪声相关的信号并将其保存在混淆变量文件中，并不对数据本身信号进行操作；而 ICA-AROMA 方法对数据的操作则更为复杂。ICA-AROMA 方法分为“激进”和“非激进”两种类型。“非激进”的 ICA-AROMA 方法对数据进行了平滑，分离并去除了数据中与头动相关的噪声，并生成了去噪后的新数据。而 ICA-AROMA “激进”方法得到的噪声相关混淆变量只保存到了混淆变量文件中，并没有在数据中去除这些噪声。在预处理过程中，研究者可以选择只使用 ComCor 而不做 ICA-AROMA，这样只得到空间标准化之后的预处理数据以及混淆变量文件；或者两种方法同时使用，这样除了可以得到空间标准化之后的预处理数据，混淆变量文件，还可以得到 ICA-AROMA 方法处理过的平滑后的预处理数据。混淆变量文件中，fmriprep 生成的混淆变量主要有三类：头动参数（主要取平移三个三个方向和转动三个方向 6 个参数），ComCor 生成的与生理噪声相关的成分信号（一般取 aComCor 的前 5 个成分即可），高通滤波函数 Cosines 相关参数。头动和 ComCor 参数主要是为了分别去除头动和生理噪声对实验任务相关信号的影响。注意 Cosines 参数是 fmriprep 使用离散余弦变换 (Discrete Cosine Transform，DCT) 的方法产生的，而使用 Cosines 参数的目的是为了去除低频噪声的干扰。因为 SPM 默认的高通滤波功能所使用的方法也是离散余弦变换 (Ashburner et al., 2014)，两种方法在原理和目的上没有差别，所以本研究中一致采用了 SPM 默认的高通滤波来去除低频噪声干扰。

chinaXiv:201908.00125v1

chinaXiv:201908.00125v1

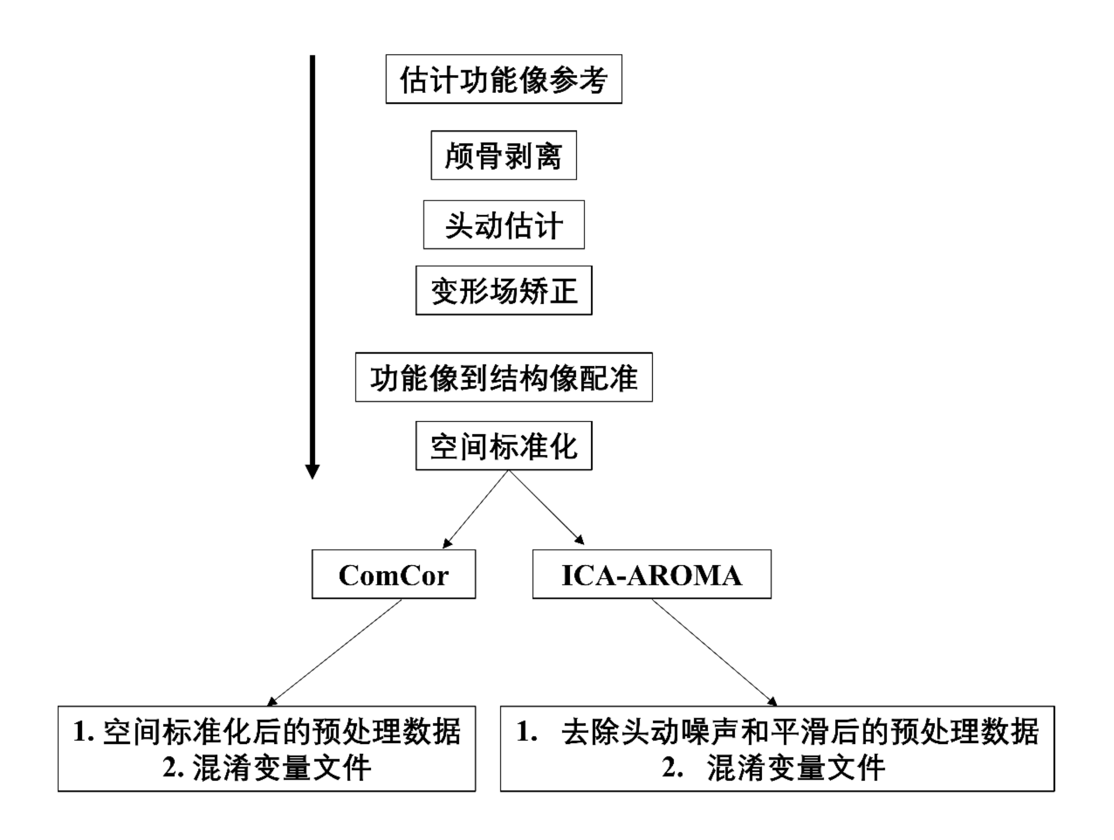


图 2. fmriprep 预处理流程简图

第三步预处理方法的选择不仅仅会产生不同的预处理结果，也影响第四步统计分析。预处理阶段方法的选择主要对第四步的统计分析带来两个问题：第一，选择哪一方法产生的预处理数据进行后续的统计分析；第二，在统计分析中需要回归掉哪些混淆变量。如果选择基于 ComCor 方法得到的空间标准化之后的预处理数据，那么在个体分析时就需要加入头动参数和与生理噪声相关的参数作为协变量；如果选择 ICA-AROMA 方法处理过的平滑后的预处理数据，那么只需要加入与生理噪声相关的参数作为协变量。另外一个需要注意的事项就是基于 ComCor 方法得到的空间标准化之后的预处理数据是没有经过平滑处理的，而 ICA-AROMA 方法处理过的数据是默认平滑过的（默认平滑核 6*6*6）。所以，对于单纯使用 ComCor 方法的分析流程，研究者还需要额外对基于 ComCor 方法得到的空间标准化之后的预处理数据进行平滑处理，再对平滑后的预处理数据进行统计分析。

新兴的 fMRI 数据整理标准 BIDS 以及相应的 BIDS APP，不仅仅试图统一数据整理的格式，也

试图统一数据处理的技术和流程。这一趋势无疑给脑成像研究者带来了极大的便利，也为脑成像数据的公开和共享铺平了道路。但是，这些新兴的技术和现有的传统脑成像数据处理软件之间的整合还并不完善。正如本文所诉，不同的预处理技术的选择决定了后续统计方法的搭配。目前并没有一个绝对的标准 fMRI 数据预处理和统计流程。研究者在使用这些新兴软件和技术时，需要根据自己的实际研究情况，考虑如何最优化脑成像数据预处理和统计流程的问题。最后，本文提供了整个预处理和个体统计分析流程的参考代码（https://github.com/ChenSD/BIDS_fmriprep），希望能够为后来者提供一点便利。

chinaXiv:201908.00125v1

参考文献

1 Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., . . . Moran, R. (2014). SPM12 manual. *Wellcome Trust Centre for Neuroimaging, London, UK*.

2 Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage, 37*(1), 90-101. doi:10.1016/j.neuroimage.2007.04.042

3 Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PloS One, 12*(9), e0184661.

4 Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M. M., . . . Devenyi, G. A. (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Computational Biology, 13*(3), e1005209.

5 Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., . . . Halchenko, Y. O. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data, 3*, 160044.

6 Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2013). Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to Carp. *NeuroImage, 76*.

7 Pruim, R. H., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage, 112*, 267-277.

chinaXiv:201908.00125v1

chinaXiv:201908.00125v1

The standardization process of organizing brain imaging data with BIDS and discussion of the process pipeline by BIDS APPs

CHEN ShengDong; GAO Wei; LUO Li; YANG JieMin; YUAN JiaJin

(*Key Laboratory of Cognition and Personality of Ministry of Education, School of Psychology, Southwest University, Chongqing* 400715, *China*)

Abstract The noninvasive functional magnetic resonance imaging (fMRI) technique has been a crucial method for brain imaging research. However, the divergence existing in the plethora of datasets acquired in the labs around the world limited the rate of progress of brain research. To address this issue, international brain scientists jointly proposed a framework named with the Brain Imaging Data Structure (BIDS) for organizing and describing neuroimaging data, and developed BIDS APPs for analyzing neuroimaging data organized in compliance with the BIDS. The present paper briefly introduced BIDS and the fMRI processing pipeline with BIDS APPs. Moreover, we discussed how to reasonably integrate the preprocessing pipeline with BIDS APPs (MRIQC and fmriprep) and the following statistical analysis. Specifically, we suggested that brain researchers should first organize their fMRI data in keeping with BIDS, and then use MRIQC APP to do automated quality control for both anatomical and functional MRI data organized in compliance with the BIDS. After eliminating the "bad data" (e.g., FD>0.2 mm), fmriprep can then be used to preprocess the "good data." The fmriprep APP introduced two new preprocessing methods (ComCor and ICA-AROMA), which have been suggested to be effective in increasing sensitivity to group-level activation. However, it should be noted that the choice of fmriprep preprocessing methods determine which regressors should be included in the following individual general linear model (GLM). If only ComCor method was used, both the head motion related-noise (6 motion parameters) and cardiac and respiratory related-noise should be used as nuisance regressors in the following GLM analysis. If both ComCor and ICA-AROMA method were used, only regress cardiac and respiratory related-noise but not the head motion related-noise should be included in the following GLM analysis. This is because that ICA-AROMA method has removed the noise related to head motion in the preprocessed data, whereas ComCor did not. Therefore, regressing

out motion-related variables in the GLM may reintroduce motion artifacts. We suggest that in the face of the emerging BIDS and BIDS APPs, brain researchers need to concern how to combine it with the traditional statistical software optimistically in order to attain a better statistical power.

Keywords: brain imaging, fMRI, data analysis pipeline, BIDS, BIDS APP

chinaXiv:201908.00125v1